Received: 8 December 2023 | Revised: 16 July 2024 | Accepted: 19 July 2024

DOI: 10.1002/asl.1265

RESEARCH ARTICLE

Atmospheric Science Letters RMetS

A hypothesis on ergodicity and the signal-to-noise paradox

Daniel J. Brener

The Higgs Centre for Theoretical Physics, The University of Edinburgh, Edinburgh, UK

Correspondence
Daniel J. Brener, The Higgs Centre for Theoretical Physics, The School of Physics & Astronomy, The University of Edinburgh, Edinburgh, EH9 3FD, UK.
Email: daniel.brener@ed.ac.uk

Funding information
Science and Technology Facilities Council, Grant/Award Number: ST/W507441/1

Abstract

This letter raises the possibility that ergodicity concerns might have some bearing on the signal-to-noise paradox. This is explored by applying the ergodic theorem to the theory behind ensemble weather forecasting and the ensemble mean. Using the ensemble mean as our best forecast of observations amounts to interpreting it as the most likely phase-space trajectory, which relies on the ergodic theorem. This can fail for ensemble forecasting systems if members are not perfectly exchangeable with each other, the averaging window is too short and/or there are too few members. We argue these failures can occur in cases such as the winter North Atlantic Oscillation (NAO) forecasts due to intransitivity or regime behaviour for regions such as the North Atlantic and Arctic. This behaviour, where different ensemble members may become stuck in different relatively persistent flow states (intransitivity) or multi-modality (regime behaviour), can in certain situations break the ergodic theorem. The problem of non-ergodic systems and models in the case of weather forecasting is discussed, as are potential mitigation methods and metrics for ergodicity in ensemble systems.

KEYWORDS

ensembles, ergodicity, NAO, signal-to-noise paradox, seasonal forecasting

1 | INTRODUCTION

The equations governing the atmosphere were shown in the seminal papers of Lorenz (1963) and Lorenz (1969) to exhibit sensitive dependence on initial conditions, commonly known as chaos. Given the impossibility of perfectly knowing these conditions, predictability of the atmosphere is inherently limited, even with a perfect model. To address this uncertainty, the method of ensemble weather prediction was developed, see Epstein (1969). Using multiple realizations of a numerical model with tiny variations in initial conditions, these small perturbations give rise to divergent predictions due to chaos. The average value of an observable over several ensemble members can sometimes be interpreted as the best estimate of the said state of the observable at or over a particular time. The mean and standard deviation of predictions across ensemble members provide estimates of an observable's state and uncertainty, despite the deterministic equations underlying the models, as discussed in Murphy and Palmer (1986) and Palmer and Hagedorn (2006).

Scaife and Smith (2018) reviewed what has come to be known as a signal-to-noise paradox: that atmosphere–

Abbreviations: GloSea5, Met Office Global Seasonal Forecast System, version 5; NAO, North Atlantic Oscillation; PDF, probability density function; RPC, ratio of predictable components.

ocean coupled climate and long range prediction ensemble models are better at predicting reality, than they are at predicting themselves. This was first raised as possibility by Kumar (2009) and consequently demonstrated by Eade et al. (2014). They derived a statistic for comparing ensemble models to observations known as the "ratio of predictable components" (RPC) for several atmospheric parameters,

$$\mathrm{RPC}^2 = \frac{r^2_{\overline{\mathrm{E}}\mathrm{O}}}{r^2_{\overline{\mathrm{E}}\mathrm{E}_i}}, \tag{1}$$

where $r_{\overline{\mathrm{E}}\mathrm{O}}$ is the Pearson correlation between the model ensemble mean and the observations, and $r_{\overline{\mathrm{E}}\mathrm{E}_i}$ is the average correlation between the model ensemble mean and a single ensemble member. It was found that the correlation between the ensemble mean and observations is often much greater than the average correlation between the ensemble mean and a single ensemble member ($\mathrm{RPC} > 1$), referred to as an anomalous RPC. Hence, a signal-to-noise paradox - the model predicts reality better than it predicts itself. This has since been reported by many different groups such as Scaife et al. (2014) Stockdale et al. (2015), Charlton-Perez et al. (2019), Weisheimer et al. (2019) and Dunstone et al. (2023).

Over the past few years, a great deal of effort from the community has been put into finding different resolutions to the paradox, and different scenarios where it arises. Strommen et al. (2023) showed that for the North Atlantic Oscillation (NAO), the paradox may be interpreted as a probabilistically under-confident forecast for occurrences of high magnitude NAO. A possible explanation for this is that the model has reduced persistence of particular atmospheric regimes, especially in the Northern Hemisphere according to Strommen and Palmer (2018) and Strommen (2020). In terms of model dynamics, weak atmosphere eddy feedback was suggested by Scaife et al. (2019) and Hardiman et al. (2022), or weak ocean–atmosphere coupling in models by Ossó et al. (2020). Early approaches to the paradox hoped that enhancing model physics could directly improve the correlation between the model and the ensemble mean, thereby increasing forecast skill. These improvements have not yet been achieved. A detailed summary of the current state of understanding and avenues to tackle the problem can be found in Weisheimer et al. (2024). Taking a different angle, Bröcker et al. (2023) argued that in some cases, that the signal-to-noise paradox should not be considered paradoxical due to the assumption that the forecast error is related to the correlation of the ensemble mean with the observations, which is not necessarily always the case. It is in this more statistical direction that this Letter takes us.

We suggest that part of the paradox could be due to a violation of the ergodic theorem which is relied upon for the statistical moments comprising the RPC. This is because using the ensemble mean as our best forecast of observations amounts to interpreting it as the most likely phase-space trajectory, which relies on the ergodic theorem. We argue that this can fail in cases such as the winter NAO forecasts due to intransitivity/multi-modality which can in certain situations break the ergodic theorem. The ideas presented are not intended to resolve the paradox entirely as there are other contributing issues related to both modelling and the definitions used in the computation of the RPC (e.g. Bröcker et al. (2023), Zhang and Kirtman (2019), Knight et al. (2022) and O'Reilly et al. (2019)). We hope this article will act as a catalyst for others to consider how the ergodicity assumption might be tested more rigorously, and serve as a reminder of it.

2 | ERGODIC THEORY

2.1 | The ergodic theorem

Consider a dynamical system, x_t evolving with time t on some phase space $X = \mathbb{R}^d$, where d is the system dimension. Suppose there exists some observable $f(x_t) \in X$, which can be integrated such that a measure μ on X is preserved. Let T be a unique time-evolution operator such that an initial state x_0 can evolve to a discrete later time x_l via $T^l x_0 = x_l,\ l \in \mathbb{N}$. If T preserves the invariant measure μ, then T is said to be ergodic. Our observable f has a *time mean of a single path* on X going from an initial position (x_0, t_0) to a later position (x_l, t_l) given by,

$$\overline{f}(x_0) = \lim_{l \to \infty} \frac{1}{l} \sum_{k=0}^{l-1} f\left(T^k x_0\right). \tag{2}$$

For the same observable f, one can also find the *space mean*,

$$\overline{f}(x_{t'}) = \int \mu f(x_{t'}) dx \tag{3}$$

at an instant in time t'. Under certain conditions on the system, the ergodic theorem holds that: the time and space means are the same for all possible initial conditions x_0. This is also known as the ergodic hypothesis and was proved, along with the existence of the averages individually, by Birkhoff (1931). See Ollagnier (1985) for a

formal approach to the subject. In the next section, we will see how this works for ensemble models.

2.2 | Application to ensemble modelling

In numerical weather prediction, x_t is the state of the atmosphere at time t, evolving in some phase space R^d, where d is very large. The operator T represents the evolution of the atmosphere forward in time according to all the equations of physics. The phase space of the various atmospheric/climatic states is some strange attractor, $\mathcal{A}$ see Tsonis and Elsner (1989) for an introduction. T takes us between states on $\mathcal{A}$, whilst preserving $\mathcal{A}$. If T is chosen to be ergodic, then T will visit all parts of $\mathcal{A}$ over time; for a rigorous treatment see Eckmann and Ruelle (1985) part 2E.

The ensemble mean of an observable, $E(f(x_t))$ at a time t is

$$E(f(x_t)) = \frac{1}{n}\sum_{i=1}^{n} f^i(x_t) \tag{4}$$

where $f^i(x_t)$ is the observable from the i^{th} member of the ensemble of size n. The significance of the ensemble size, n has been well studied on all temporal and spatial scales; see Palmer and Hagedorn (2006) and Leutbecher (2019). Under a "signal + noise" model, the idea is that forecast skill grows with n due to the suppression of unpredictable noise.

In using the RPC metric, we want to interpret the ensemble mean as being the *most likely trajectory* in the phase space X, and hence our best guess for what the observations will do. This is an ergodicity assumption as we want to equate the space average (ensemble mean) with the most likely trajectory (time average). An ensemble is initialised as a set of perturbed initial states $\{x_0^i\}$. There must exist a unique time-evolution operator T_i^k for each member that evolves each initial state over a time $\Delta T \equiv T_{k=l} - T_{k=0}$. So our ensemble mean can be given by,

$$E(f(x_l)) = \frac{1}{n \cdot l}\sum_{i=1}^{n}\sum_{k=0}^{l-1} f\left(T_i^k x_0^i\right) \tag{5}$$

In other words, running our ensemble simulation is equivalent to computing the evolution of the initial states by the operator T. Since T is unique, it is more convenient to think of each ensemble member as having their own continuous time-dependant probability density function (PDF), $P^i(x_t)$, that evolves according to unique T_i^l, so $T_i^l P^i(x_0) = P^i(x_l)$. This gives the ensemble mean as

$$E(f(x_t)) = \frac{1}{n \cdot \Delta T}\int_{T_0}^{T} f^1(x_t)P^1(x_t) + f^2(x_t)P^2(x_t) + \ldots + f^n(x_t)P^n(x_t)dt, \tag{6}$$

so that the observable as measured in each member is weighted by the evolving PDF unique to each member. This average will only represent the most likely trajectory if the following conditions hold:

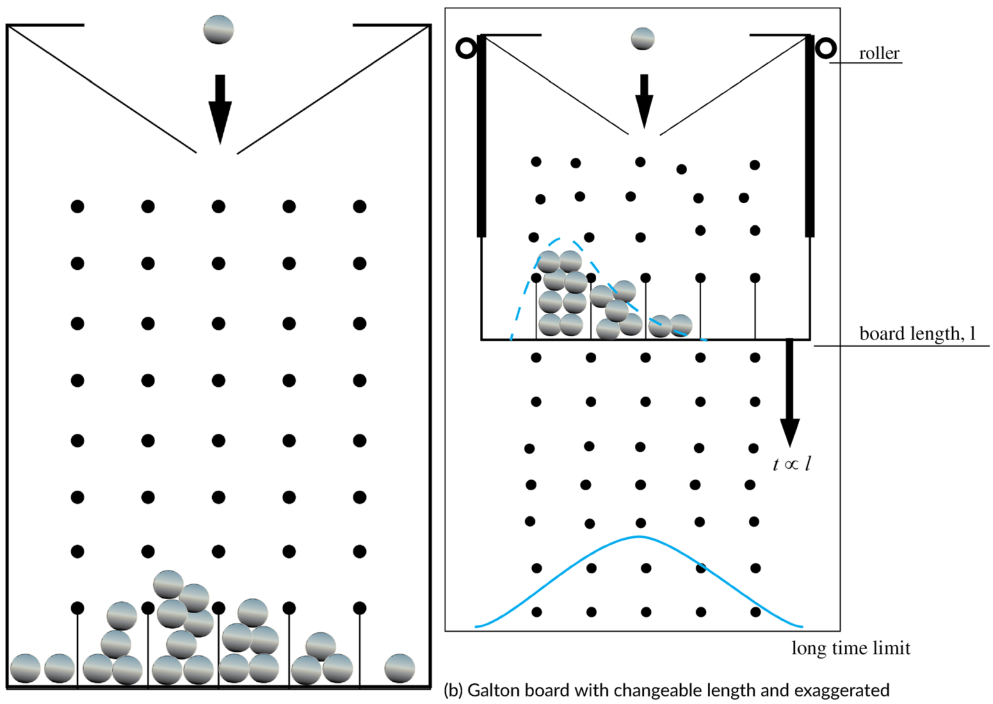
$$\begin{aligned}
&\lim n \to \infty; \text{infinite ensemble size},\\
&\lim \Delta T \to \infty; \text{infinitely long averaging window},\\
&\left.\frac{dP^i}{dt}\right|_{t=\Delta T,\forall i} = 0; \text{distribution function for each ensemble}\\
&\qquad \text{member is stationary as } T^i \text{ must be}\\
&\qquad \text{unique over period } \Delta T.
\end{aligned}$$

The final condition is equivalent to saying that the ensemble members must be *exchangeable*, which is a common assumption for ensemble prediction systems and will be discussed in a later section. We will next interrogate these concepts using a Galton Board for Gedankenexperimente.

2.3 | Ergodicity in the Galton board: Winding a classic model

A Galton board is a device invented by Galton (1889) that consists of a vertical board with an array of pegs. At the top, there is an entry point from which small balls, typically marbles, are released. The Galton board is often used to visually illustrate how randomness, combined with a large number of trials (balls), leads to a Gaussian distribution. In meteorology, it has been used to teach the principles of ensemble forecasting and has also been used to develop conceptual arguments in this field. We illustrate a normal Galton board in Figure 1a where after multiple individual balls we begin to approximate a Gaussian.

For our purposes, each individual ball trajectory represents the time evolution of a single ensemble member and the statistical distribution of the final positions represents the ensemble average for the model. As one observes more and more balls passing through the board, the distribution of final positions at the bottom converges to a pattern that resembles a Gaussian distribution. This convergence is analogous to the ergodic theorem, where the time average (individual ball's trajectory) converges to the ensemble average (statistical distribution of final positions) over a large number of trials. If we imagine an *infinitely long board*, the long-time average trajectory for



(a) Fixed length classic Galton board.

(b) Galton board with changeable length and exaggerated peg-misalignment.

FIGURE 1 Illustration of our two Galton boards. Not drawn to scale. The right-hand board has adjustable length using a roller. Adjusting the length corresponds to adjusting the time the balls have to explore the system. The longer the board, the longer the time the balls spend in the system and the better representative the ensemble mean and variance will be of the underlying probability density function (PDF), even if the space is multi-modal. Blue solid line in (b) indicates the Gaussian outcome of the long-board (time) limit, in contrast to distribution of balls (blue dashed) with some fixed short board length (time). Note that the distribution of balls is not instantaneous, but an average over a number of different balls (ensemble members) over a time proportional to the length they are allowed to drop.

a single ball will be straight down; equivalent to the position of the ensemble mean. It is also worth noting from Kumičák (2000) and Hoover and Moran (1992) that the classic Galton board has been found to possess an ergodic strange attractor, and from Judd (2007) that it is not simply a random-walk phenomena. Hence, it has useful parallels to weather and climate prediction.

We can play with this idealised ensemble model to examine what can happen if the ergodic assumption fails. The way this can happen is if the system contains multimodality, in other words, there exist potential wells via certain trajectories into which the balls can stumble from which they are less likely to recover in a certain timeframe. This does not have to be dependent on the different initial conditions. In simulations of Galton boards such potential wells have been found to exist, see Ahmed et al. (2022).

Consider a Galton board of *finite length* where we are able to modify this length using a roller, thus change the length of time that the balls must spend in the system. Now let us suppose that the gaps between the pegs are not identically uniform, so at certain points the balls path is impeded slightly more. Locally these act as potential wells, creating multi-modal features. The longer we make the board, the more time the balls will have to explore, enter and escape these potentials. With a sufficiently long board (time), the resulting distribution, on average, will be our Gaussian again. However, the shorter the board, the more the multi-modal features dominate the picture. One can imagine all kinds of situations that could occur, such as where the balls get stuck in roughly the same region of the board and so do not explore all the possible trajectories adequately. The longer we make the board, hence the longer the time, the more Gaussian our distribution of balls will become. We also need to use enough balls (large enough ensemble) to explore the board adequately. The key point is that the underlying distributions for the space/ensemble and time averages will be different when ergodicity fails.

Our model can also be extended to incorporate external forcings. For instance, a light fan blowing across the board would shift the distributions, similar to the conceptual model of Palmer (1999). The main difference between our Galton board and the picture of Palmer (1999) is that the latter constrains the system to a bimodal structure which is examined in the long time-

limit, whereas our system allows us to imagine any number of modes (depending on the board "resolution" and complexity) and our roller lets us examine long and short time limits. One can also think about exchangeability, as if one were to drop different kinds of balls (e.g. size/mass/stickiness), the final distributions would not make sense as a predictor for any one particular ball.

We can now see that without enough balls and or if the averaging time is too short, the ergodic theorem is not satisfied. In the case of the NAO, which evolves roughly daily, then for a 90-day season in which the flow exhibits multimodality, 90 days will not always be sufficient to explore the complete phase space. This is because such ensemble statistics reflect only those states the system has visited, highlighting the impact of daily variations on the dynamics that govern long-term averages.

3 | FAILURE OF ERGODICITY AND INTERPRETATION OF THE RPC

We will now examine how failure of the ergodic assumption could lead to anomalous RPC values and hence the signal-to-noise paradox. To do this, we draw on the statistical framework developed by Bröcker et al. (2023). Let $V(e)$ and $V(o)$ be the respective variances of the ensemble and the observations, then it has been shown that,

$$\sqrt{\frac{V(e)}{V(o)}} < r_{\overline{\mathrm{E}}\mathrm{O}} \Rightarrow \mathrm{RPC} > 1, \tag{7}$$

where $r_{\overline{\mathrm{E}}\mathrm{O}}$ is the Pearson correlation between the observations and the ensemble mean. Bröcker et al. (2023) showed that when we have situations of small correlations, something highly likely for seasonal forecasting, then even minor differences between the observations and ensemble variance gives rise to a signal-to-noise paradox via the above inequality.

We make the conjecture that failure of ergodicity in an ensemble can in some cases lead to these differences. If the ergodic assumption fails for the mean, as defined by Equation 6, then it also fails for *all* statistical moments, including the variance $V(e)$. Mathematically, this is a strong statement as it renders such a variance as meaningless, as without ergodicity, the distribution will be different from sampling either from a long time-series of a single member or from a complete ensemble. Like in our Galton board Figure 1b, the long-time distribution is a Gaussian, but with too short a board we are more likely to get some skewed distribution. The magnitude of the difference is something which can only be determined by direct computation of the distributions. However, simply failing to respect ergodicity does not mean that $V(e)$ should systematically underestimate $V(o)$.

Let us try a thought experiment with our windable Galton board in Figure 1b. Suppose the observational reality is that the ball falls straight down, round a zero-mean Gaussian. We could conceive of a situation where our ensemble of balls gets stuck in a potential well such that they clump together relatively close to the centre. This would result in $V(e) < V(o)$ with a large $r_{\overline{\mathrm{E}}\mathrm{O}}$, hence an anomalous RPC value. From the weather regime perspective of the NAO, if one has multiple persistent regimes then a single 90 day season will not always be sufficient for a finite ensemble to adequately sample the space, necessary for the ergodic theorem. The members will appear to become "stuck" in certain regimes. Consequently, each member on average then underestimates the total variance of a given season, leading to $V(e) < V(o)$.

The spatial inhomogeneity of the anomalous RPC values (see e.g. Weisheimer et al. (2019) for detailed study), especially over the North Atlantic and Arctic as reported by Cottrell et al. (2024) strongly suggest multimodality is involved, as these regions correlate with well-studied multi-modal behaviour. Strommen and Palmer (2018) point out that a multi-modal or regime approach provides an alternative perspective on the paradox. Using a toy bimodal Markov-chain model, they obtain anomalous RPC values when the persistence of the modes is underrepresented. They note that the potential wells for the regimes may be too shallow in the models. Additionally, Falkena et al. (2022) found that using more regimes improved the representation of wintertime Euro-Atlantic sector dynamics. We posit that this multimodality, in breaking the ergodic assumption necessary for the ensemble mean and variance, would consequently invalidate interpretation of those statistics, such as RPC. In fact, it leads one to reach the same conclusion as Bröcker et al. (2023), the signal-to-noise "paradox" is not really paradoxical; because ergodically we cannot expect the ensemble mean/variance to be the most likely phase-space trajectory.

Whilst our approach using the ergodic theorem to ensembles is new to this problem, we note that Lorenz discussed transitive and intransitive systems and climate, see Lorenz (1968), Lorenz (1976) and Lorenz (1990). Ergodicity and transitivity are essentially the same mathematical phenomenon (see pedagogical discussion by Shalizi (2007)). Lorenz's discussion of transitivity sets up in more general terms the concepts for climatic regime states which were later more clearly defined by Palmer (1999). The ergodic theorem therefore provides further supporting rigour and a different perspective on the multi-modality paradigm.

One mitigation strategy for the multi-modality issue is to increase the ensemble size. However, even with a relatively large ensemble/sample, the paradox has been found to persist, for example see Cottrell et al. (2024) and Shi et al. (2015). There is weak evidence that the paradox exists on longer than multi-decadal timescales Scaife and Smith (2018). The reason for this may be understood from an ergodic perspective, as to define a mean climate state the interval ΔT over which we compute the average, is sufficiently long that the distributions P^i converge Tantet (2016). One could increase the averaging window to see how this affects the RPC. However one cannot increase it too much else the definition of the observable would be changed. In other words, for very large averaging windows, the observable simply becomes a different kind of climatic average: sub-seasonal to seasonal, seasonal to annual, etc.

Another way in which failure of ergodicity can occur is if the ensemble members are not exchangeable with each other. This can be seen from Equation 6, as one cannot factorise a single $P^i(x_t)$ in the integrand unless they are all *perfectly* exchangeable. According to Leutbecher (2019), the ensemble members of the operational ECMWF model, which has reported anomalous RPCs, are not truly exchangeable because its initial perturbations have a plus-minus symmetry (i.e., initial perturbation of member $2k$ is minus the perturbation of member $(2k-1)$), introducing systematic differences between ensemble members. Other forecasting systems which are known to not satisfy exchangeability include the Meteo-France global ensemble, and Canadian ensemble prediction system, see Leutbecher et al. (2017). Also note that Siegert et al. (2016), see section 4b, raised the possibility that the Met Office's GloSea5 climate prediction system ensemble members might not be exactly exchangeable. As reported by Weisheimer et al. (2024), it is understood that exchangeability of observations and ensemble members provides a strong criterion for the signal-and-noise paradox. What is not recognised explicitly is that the members themselves must also be exchangeable to avoid failure of the ergodic theorem when interpreting the ensemble mean/variance as the most likely trajectory and then using these for the RPC. Combined with multi-modality, non-exchangeability could make it more likely for situations of $V(e) < V(o)$ to occur as we need to correctly weight each member according to its uniquely evolved PDF. Members might appear "stuck" - in contrast to their evolved PDFs, which if used as weights would make the ensemble mean and variance better estimates of the most likely phase-space trajectory.

If we look at other simpler physical systems we readily find cases where failure of the ergodic assumption leads to significant incorrect statistical measures. Examples of such systems include supercooled liquids to glass transitions (see Thirumalai et al. (1989)), diffusive processes within the plasma membrane of living cells (see Weigel et al. (2011)) or general noise processes (see Mangalam and Kelty-Stephen (2022)). In each of these cases, handling ensemble averages without due attention to the ergodic assumption leads to in simple terms, wrong answers. Peters and Klein (2013) demonstrates an example of this in geometric brownian motion where nonergodicity can lead to the ensemble mean growing exponentially, whilst simultaneously any individual trajectory decays exponentially according to its time average.

4 | SUGGESTIONS FOR FUTURE STUDIES

Future work to examine the hypotheses in this Letter could include testing the Thirumalai–Mountain effective ergodic convergence metric, $\Omega(t)$ measures the effective ergodicity by the difference between the time average of an observable and its ensemble average over the entire system. So for an ensemble of size n,

$$\Omega(t) = \frac{1}{n}\sum_{i=1}^{n}\left[f(t)_i - E(f(t))\right]^2, \tag{8}$$

where the ensemble average is as defined by Equation 4. In the long-time limit, $\Omega \to 0$ for an ergodic system. The rate at which this occurs can be used to indicate timescales over which an ergodic approximation might be appropriate. This metric has, as far as we are aware, not been tested before in ensemble prediction systems. However, it is utilised for earthquake forecasting see for example Tiampo et al. (2003) and Tiampo et al. (2007), and studies of fluids, see de Souza and Wales (2005). For other metrics related to ergodicity, see Mathew and Mezić (2011).

Given sufficient computing resources, one could develop an "ensemble-of-ensembles" approach to test perfect exchangeability. Once could generate at each time step small perturbations to each member generating an ensemble per member, so as to compute finite time Lyapunov exponents. From this, one could compute the Lyapunov spectrum per main member. If the spectrum varies between the ensemble members then this would imply that fundamentally the members evolve into different state spaces, and thus are not exchangeable, hence failure of the ergodic mean to represent the most likely trajectory.

An approach to relax the ergodicity constraint but still make use of ensembles is to enforce dynamical invariants. These are quantities that are conserved along the phase-space path of the system. For example, in the probability theory for non-equilibrium gravitational systems developed by Peñarrubia (2015), dynamical invariants are used with the explicit purpose of not relying ergodicity assumptions. For turbulent fluids, the most obvious dynamical invariant would be the Lyapunov spectrum and fractal dimension, besides vorticity. Recently Platt et al. (2023) developed such a scheme for a numerical weather model machine learning training method. They tested their model on the classic Lorenz 1996 chaotic dynamical system, and found including ergodic constraints improved forecast skill.

5 | SUMMARY

The signal-to-noise paradox, assessed by the RPC metric, relies on interpreting ensemble mean and variance as the most probable phase-space trajectory. However, this interpretation relies on the ergodic theorem, which faces potential failure due to insufficient ensemble size, short averaging windows, and non-exchangeability of members. In the cases like winter NAO forecasts, evidence of multi-modal regime behaviour or intransitivity between NAO phases, irrespective of multi-modality, could exacerbate this failure, leading to hindered exploration of the complete phase space. Consequently, the ensemble mean and variance cannot reliably represent the most likely trajectory, resulting in the paradox when the ensemble becomes stuck – correlating with observations but underestimating the variance.

AUTHOR CONTRIBUTION

Daniel J. Brener: Conceptualization; formal analysis; investigation; methodology; visualization; writing – original draft preparation, writing – review & editing; project administration.

ACKNOWLEDGEMENTS

D.J.B. was funded by the UK Science and Technology Facilities Council. He was particularly grateful to Professor Adam Scaife who introduced him to the problem, as well as Dr Jose Rodriguez and Dr Leon Hermanson of the Met Office and Professor Jorge Peñarrubia of The University of Edinburgh, for giving their time to insightful tête-à-tête discussions. He also thanks Alexander Tully of The University of Liverpool for comments on the manuscript. D.J.B. expresses his gratitude to two anonymous reviewers and Associate Editor Dr. Steven Hardiman for their invaluable comments and feedback.

FUNDING INFORMATION

Science and Technology Facilities Council, UK.

CONFLICT OF INTEREST STATEMENT

The author declares no conflict of interest.

DATA AVAILABILITY STATEMENT

No data was used or produced in the course of this research.

ORCID

Daniel J. Brener https://orcid.org/0000-0003-3340-0177

REFERENCES

Ahmed, J., Chumley, T., Cook, S., Cox, C., Grant, H., Petela, N. et al. (2022) Dynamics of the no-slip Galton board. Available at: https://arxiv.org/abs/2208.07790

Birkhoff, G.D. (1931) Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences of the United States of America*, 17, 656–660 Available at: http://www.jstor.org/stable/86016

Bröcker, J., Charlton–Perez, A.J. & Weisheimer, A. (2023) A statistical perspective on the signal-to-noise paradox. *Quarterly Journal of the Royal Meteorological Society*, 149, 911–923. Available from: https://doi.org/10.1002/qj.4440

Charlton-Perez, A., Bröcker, J., Stockdale, T.N. & Johnson, S.J. (2019) When and where do ECMWF seasonal forecast systems exhibit anomalously low signal-to-noise ratio? *Quarterly Journal of the Royal Meteorological Society*, 145, 3466–3478. Available from: https://doi.org/10.1002/qj.3631

Cottrell, F.M., Screen, J.A. & Scaife, A.A. (2024) Signal-to-noise errors in free-running atmospheric simulations and their dependence on model resolution. *Atmospheric Science Letters*, 25, e1212. Available from: https://doi.org/10.1002/asl.1212

de Souza, V.K. & Wales, D.J. (2005) Diagnosing broken ergodicity using an energy fluctuation metric. *The Journal of Chemical Physics*, 123, 134504. Available from: https://doi.org/10.1063/1.2035080

Dunstone, N., Smith, D.M., Hardiman, S.C., Hermanson, L., Ineson, S., Kay, G. et al. (2023) Skilful predictions of the summer north Atlantic oscillation. *Communications Earth & Environment*, 4, 409. Available from: https://doi.org/10.1038/s43247-023-01063-2

Eade, R., Smith, D., Scaife, A.A., Wallace, E., Dunstone, N., Hermanson, L. et al. (2014) Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, 41, 5620–5628. Available from: https://doi.org/10.1002/2014gl061146

Eckmann, J.P. & Ruelle, D. (1985) Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics*, 57, 617–656. Available from: https://doi.org/10.1103/RevModPhys.57.617

Epstein, E.S. (1969) Stochastic dynamic prediction. *Tellus*, 21, 739–759.

Falkena, S.K., de Wiljes, J., Weisheimer, A. & Shepherd, T.G. (2022) Detection of interannual ensemble forecast signals over the North Atlantic and Europe using atmospheric circulation regimes. *Quarterly Journal of the Royal Meteorological Society*, 148, 434–453. Available from: https://doi.org/10.1002/qj.4213

Galton, F. (1889) *Natural inheritance*. No. v. 42; v. 590. London: Macmillan.

Hardiman, S.C., Dunstone, N.J., Scaife, A.A., Smith, D.M., Comer, R., Nie, Y. et al. (2022) Missing eddy feedback may explain weak signal-to-noise ratios in climate predictions. *NPJ Climate and Atmospheric Science*, 5, 57.

Hoover, W.G. & Moran, B. (1992) Viscous attractor for the Galton board. *Chaos*, 2, 599–602.

Judd, K. (2007) Galton's quincunx: random walk or chaos? *International Journal of Bifurcation and Chaos*, 17, 4463–4469.

Knight, J.R., Scaife, A.A. & Maidens, A. (2022) An extratropical contribution to the signal-to-noise paradox in seasonal climate prediction. *Geophysical Research Letters*, 49, e2022GL100471. Available from: https://doi.org/10.1029/2022GL100471

Kumar, A. (2009) Finite samples and uncertainty estimates for skill measures for seasonal prediction. *Monthly Weather Review*, 137, 2622–2631.

Kumičák, J. (2000) Galton board as a model for fluctuations. In *Unsolved Problems of Noise and Fluctuations: UPoN'99: Second International Conference*, vol. 511 of *American Institute of Physics Conference Series*, 144–149.

Leutbecher, M. (2019) Ensemble size: how suboptimal is less than infinity? *Quarterly Journal of the Royal Meteorological Society*, 145, 107–128. Available from: https://doi.org/10.1002/qj.3387

Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S.T.K., Balsamo, G., Bechtold, P. et al. (2017) Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143, 2315–2339. Available from: https://doi.org/10.1002/qj.3094

Lorenz, E.N. (1963) Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20, 130–141.

Lorenz, E.N. (1968) *Climatic determinism*. Boston, MA: American Meteorological Society, pp. 1–3. Available from: https://doi.org/10.1007/978-1-935704-38-6_1

Lorenz, E.N. (1969) The predictability of a flow which possesses many scales of motion. *Tellus*, 21, 289–307.

Lorenz, E.N. (1976) Nondeterministic theories of climatic change. *Quaternary Research*, 6, 495–506 Available at: https://www.sciencedirect.com/science/article/pii/0033589476900223

Lorenz, E.N. (1990) Can chaos and intransitivity lead to interannual variability? *Tellus A*, 42, 378–389. Available from: https://doi.org/10.1034/j.1600-0870.1990.t01-2-00005.x

Mangalam, M. & Kelty-Stephen, D.G. (2022) Ergodic descriptors of non-ergodic stochastic processes. *Journal of the Royal Society Interface*, 19, 20220095. Available from: https://doi.org/10.1098/rsif.2022.0095

Mathew, G. & Mezić, I. (2011) Metrics for ergodicity and design of ergodic dynamics for multi-agent systems. *Physica D: Nonlinear Phenomena*, 240, 432–442 Available at: https://www.sciencedirect.com/science/article/pii/S016727891000285X

Murphy, J. & Palmer, T. (1986) Experimental monthly long-range forecasts for the united-kingdom. 2. A real-time long-range-forecast by an ensemble of numerical integrations. *Meteorological Magazine*, 115, 337–349.

Ollagnier, J.M. (1985) *Ergodic theory and statistical mechanics, vol. 1115 of lecture notes in mathematics*. Berlin, Heidelberg: Springer.

O'Reilly, C.H., Weisheimer, A., Woollings, T., Gray, L.J. & MacLeod, D. (2019) The importance of stratospheric initial conditions for winter north Atlantic oscillation predictability and implications for the signal-to-noise paradox. *Quarterly Journal of the Royal Meteorological Society*, 145, 131–146. Available from: https://doi.org/10.1002/qj.3413

Ossó, A., Sutton, R., Shaffrey, L. & Dong, B. (2020) Development, amplification, and decay of Atlantic/European summer weather patterns linked to spring north Atlantic sea surface temperatures. *Journal of Climate*, 33, 5939–5951.

Palmer, T. & Hagedorn, R. (2006) *Predictability of weather and climate*. Cambridge: Cambridge University Press.

Palmer, T.N. (1999) A nonlinear dynamical perspective on climate prediction. *Journal of Climate*, 12, 575–591 Available at: https://journals.ametsoc.org/view/journals/clim/12/2/1520-0442_1999_012_0575_andpoc_2.0.co_2.xml

Peñarrubia, J. (2015) A probability theory for non-equilibrium gravitational systems. *Monthly Notices of the Royal Astronomical Society*, 451, 3537–3550. Available from: https://doi.org/10.1093/mnras/stv1146

Peters, O. & Klein, W. (2013) Ergodicity breaking in geometric brownian motion. *Physical Review Letters*, 110, 100603. Available from: https://doi.org/10.1103/PhysRevLett.110.100603

Platt, J.A., Penny, S.G., Smith, T.A., Chen, T.-C. & Abarbanel, H.D.I. (2023) Constraining chaos: enforcing dynamical invariants in the training of reservoir computers. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33, 103107. Available from: https://doi.org/10.1063/5.0156999

Scaife, A.A., Arribas, A., Blockley, E.W., Brookshaw, A., Clark, R.T., Dunstone, N. et al. (2014) Skillful long-range prediction of European and north American winters. *Geophysical Research Letters*, 41, 2514–2519. Available from: https://doi.org/10.1002/2014gl059637

Scaife, A.A., Camp, J., Comer, R., Davis, P., Dunstone, N., Gordon, M. et al. (2019) Does increased atmospheric resolution improve seasonal climate predictions? *Atmospheric Science Letters*, 20, e922.

Scaife, A.A. & Smith, D. (2018) A signal-to-noise paradox in climate science. *NPJ Climate and Atmospheric Science*, 1, 28. Available from: https://doi.org/10.1038/s41612-018-0038-4

Shalizi, C. (2007) Ergodicity and metric transitivity. Available at: https://www.stat.cmu.edu/~cshalizi/754/notes/lecture-25.pdf. [Accessed 5 July 2024]

Shi, W., Schaller, N., MacLeod, D., Palmer, T.N. & Weisheimer, A. (2015) Impact of hindcast length on estimates of seasonal climate predictability. *Geophysical Research Letters*, 42, 1554–1559. Available from: https://doi.org/10.1002/2014GL062829

Siegert, S., Stephenson, D.B., Sansom, P.G., Scaife, A.A., Eade, R. & Arribas, A. (2016) A Bayesian framework for verification and recalibration of ensemble forecasts: how uncertain is NAO predictability? *Journal of Climate*, 29, 995–1012. Available from: https://doi.org/10.1175/jcli-d-15-0196.1

Stockdale, T.N., Molteni, F. & Ferranti, L. (2015) Atmospheric initial conditions and the predictability of the arctic oscillation. *Geophysical Research Letters*, 42, 1173–1179. Available from: https://doi.org/10.1002/2014gl062681

Strommen, K. (2020) Jet latitude regimes and the predictability of the north Atlantic oscillation. *Quarterly Journal of the Royal Meteorological Society*, 146, 2368–2391.

Strommen, K., MacRae, M. & Christensen, H. (2023) On the relationship between reliability diagrams and the "signal-to-noise

paradox". *Geophysical Research Letters*, 50, e2023GL103710. Available from: https://doi.org/10.1029/2023GL103710

Strommen, K. & Palmer, T. (2018) Signal and noise in regime systems: a hypothesis on the predictability of the north Atlantic oscillation. *Quarterly Journal of the Royal Meteorological Society*, 145, 147–163. Available from: https://doi.org/10.1002/qj.3414

Tantet, A.J.J. (2016) Ergodic theory of climate: variability, stability and response. Ph.D. thesis, Utrecht University.

Thirumalai, D., Mountain, R.D. & Kirkpatrick, T.R. (1989) Ergodic behavior in supercooled liquids and in glasses. *Physical Review A*, 39, 3563–3574. Available from: https://doi.org/10.1103/PhysRevA.39.3563

Tiampo, K.F., Rundle, J.B., Klein, W., Holliday, J., Sá Martins, J.S. & Ferguson, C.D. (2007) Ergodicity in natural earthquake fault networks. *Physical Review E*, 75, 66107. Available from: https://doi.org/10.1103/PhysRevE.75.066107

Tiampo, K.F., Rundle, J.B., Klein, W., Martins, J.S.S. & Ferguson, C.D. (2003) Ergodic dynamics in a natural threshold system. *Physical Review Letters*, 91, 238501. Available from: https://doi.org/10.1103/PhysRevLett.91.238501

Tsonis, A.A. & Elsner, J.B. (1989) Chaos, strange attractors, and weather. *Bulletin of the American Meteorological Society*, 70, 14–23 Available at: https://journals.ametsoc.org/view/journals/bams/70/1/1520-0477_1989_070_0014_csaaw_2_0_co_2.xml

Weigel, A.V., Simon, B., Tamkun, M.M. & Krapf, D. (2011) Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 6438–6443.

Weisheimer, A., Baker, L.H., Bröcker, J., Garfinkel, C.I., Hardiman, S.C., Hodson, D.L.R. et al. (2024) The signal-to-noise paradox in climate forecasts: revisiting our understanding and identifying future priorities. *Bulletin of the American Meteorological Society*, 105, E651–E659 Available at: https://journals.ametsoc.org/view/journals/bams/105/3/BAMS-D-24-0019.1.xml

Weisheimer, A., Decremer, D., MacLeod, D., O'Reilly, C.H., Stockdale, T.N., Johnson, S.J. et al. (2019) How confident are predictability estimates of the winter north Atlantic oscillation. *Quarterly Journal of the Royal Meteorological Society*, 145, 140–159. Available from: https://doi.org/10.1002/qj.3446

Zhang, W. & Kirtman, B.P. (2019) Understanding the signal-to-noise paradox with a simple Markov model. *Geophysical Research Letters*, 46, 13308–13317. Available from: https://doi.org/10.1029/2019gl085159t

How to cite this article: Brener, D. J. (2024). A hypothesis on ergodicity and the signal-to-noise paradox. *Atmospheric Science Letters*, e1265. https://doi.org/10.1002/asl.1265